A Class of Simple Distribution-free Rank-based Unit Root Tests (Revision of DP 2010-72)

Hallin, M.; van den Akker, R.; Werker, B.J.M.

Publication date:
2011

CentER Discussion Paper

No. 2011-002

A CLASS OF SIMPLE DISTRIBUTION-FREE RANK-BASED UNIT ROOT TESTS

By Marc Hallin, Ramon van den Akker, Bas J.M. Werker

January 2011

This is a revised version of CentER Discussion Papers

No. 2010-72
June 2010

No. 2009-02
September 2009

ISSN 0924-7815

Tilburg University

A Class of Simple Distribution-Free Rank-Based Unit Root Tests

Marc Hallin[a,b,c,*,1], Ramon van den Akker[c], Bas J.M. Werker[c,d]

[a] *ECARES, Université Libre de Bruxelles*
[b] *ORFE, Princeton University*
[c] *Econometrics group, CentER, Tilburg University*
[d] *Finance group, CentER, Tilburg University*

Abstract

We propose a class of distribution-free rank-based tests for the null hypothesis of a unit root. This class is indexed by the choice of a *reference* density g, which needs not coincide with the unknown actual innovation density f. The validity of these tests, in terms of exact finite sample size, is guaranteed, irrespective of the actual underlying density, by distribution-freeness. Those tests are locally and asymptotically optimal under a particular asymptotic scheme, for which we provide a complete analysis of asymptotic relative efficiencies. Rather than asymptotic optimality, however, we emphasize finite-sample performances. Finite-sample performances of unit root tests, however, depend quite heavily on initial values. We therefore investigate those performances as a function of initial values. It appears that our rank-based tests significantly outperform the traditional Dickey-Fuller tests, as well as the more recent procedures proposed by Elliot, Rothenberg, and Stock (1996), Ng and Perron (2001), and Elliott and Müller (2006), for a broad range of initial values and for heavy-tailed innovation densities. As such, they provide a useful complement to existing techniques.

Key words: Unit root, Dickey-Fuller test, Local Asymptotic Normality, Rank test.
JEL codes: C12, C22.

*Corresponding address: ECARES CP 114/04, Avenue Roosevelt 50, 1050 Brussels, Belgium, Tel: +32 2 650 46 03, Fax: +32 2 650 40 12.

Email addresses: `mhallin@ulb.ac.be` (Marc Hallin), `R.vdnAkker@TilburgUniversity.nl` (Ramon van den Akker), `Werker@TilburgUniversity.nl` (Bas J.M. Werker)

[1] Royal Academy of Belgium.

1. Introduction

1.1. Autoregressive unit root models

The econometric and statistical literature dealing with near unit root asymptotics in time series models is overabundant. The presence or absence of unit roots in econometric models indeed has crucial economic policy implications. Even a short review of the literature is impossible here, and we refer to Haldrup and Jansson (2006) for a recent survey.

Unit root problems generally lead to non-standard asymptotics. The study of least-squares estimators in zero-mean unit-root autoregressive processes started with White (1958), but gained attention more widely after the publication of Dickey and Fuller (1979); unit root testing problems were first studied in detail in Dickey and Fuller (1981).

In this paper, we restrict ourselves to the simplest possible case of a univariate AR(1) unit root model with i.i.d. innovations. Extensions to multivariate settings, cointegration, panel data, more elaborate trends involving covariates, and heteroskedastic innovations fall within the general ideas of the present paper but their technical implications are not pursued here. Examples of such extensions are Phillips (1987), Chan and Wei (1988), Phillips and Perron (1988), Perron (1988), West (1988), Johansen (1991), Phillips (1991), Levin, Lin and Chu (2002), Im, Pesaran, and Shin (2003), and Elliott and Jansson (2003), to name only a few.

Within that very simple context, we are interested in the construction of "efficient" tests of the null hypothesis of a unit root. Whether theoretical asymptotic optimality results or simulations are considered, assessing the "efficiency" of such tests requires embedding the null hypothesis of a unit root into a broader model of AR(1) dependence. The literature (see, for instance, the monographs by Hamilton (1994) or Enders (2004)) traditionally considers two of them, under which the observation $(Y_1, \ldots, Y_n)$ either is generated from

– Model (a) (a very simple model of the ARMAX type[2])

$$Y_t = \rho Y_{t-1} + \mu + \varepsilon_t, \tag{1}$$

[2]Hamilton (1994) and Enders (2004) actually consider a slightly more general equation, of the form $Y_t = \rho Y_{t-1} + \mu + \gamma t + \varepsilon_t$; see Remark 2.3.

or from

– Model (b) (the so-called *components model*)

$$(Y_t - m) = \rho(Y_{t-1} - m) + \varepsilon_t. \tag{2}$$

In both cases, it is generally assumed that $\{\varepsilon_t,\ t \in \mathbb{N}\}$ is an i.i.d. innovation process, with mean zero and variance σ_ε^2, and a distribution function F admitting a density f. As for the initial value Y_0, it is often assumed to be equal to zero in Model (a), or to the stationary mean m in Model (b). It is safer, however, to leave the distribution P^{Y_0} of Y_0 unspecified, provided that Y_0 and the ε_t's are mutually independent, and that P^{Y_0} does not depend on the parameters ρ, μ or m; Y_0 then is *ancillary*, and inference on ρ is naturally conducted conditionally on Y_0.

Intuitively, Model (a) describes an autoregressive scheme in which the random shocks are i.i.d. with constant mean μ, whereas in Model (b) the i.i.d. shocks have mean zero, while the observations have (constant) mean m.

For $\rho < 1$, those two models, under two parameterizations, actually strictly coincide: indeed, (1) and (2), for $\mu = (1-\rho)m$, describe the same autoregressive data-generating process. As for $\rho = 1$, Model (a) takes the form

$$H_0: \qquad Y_t - Y_{t-1} = \mu + \varepsilon_t, \qquad \mu \in \mathbb{R} \text{ unspecified} \tag{3}$$

yielding the (first- as well as second-order nonstationary) random walk

$$Y_t = Y_0 + \mu t + u_t, \qquad u_t := \sum_{s=1}^{t} \varepsilon_s \tag{4}$$

with conditional drift $\mathrm{E}[Y_t|Y_0] = Y_0 + \mu t$ and conditional variance $\mathrm{Var}(Y_t|Y_0) = t\sigma_\varepsilon^2$. That null hypothesis H_0 strictly contains the null hypothesis

$$H_0^{(b)}: \qquad Y_t - Y_{t-1} = \varepsilon_t, \qquad m \in \mathbb{R} \text{ unspecified} \tag{5}$$

(m under $H_0^{(b)}$ is not identified) induced by Model (b), which characterizes the second-order nonstationary but first-order stationary random walk

$$Y_t = Y_0 + u_t, \qquad u_t := \sum_{s=1}^{t} \varepsilon_s \tag{6}$$

with constant conditional mean $\mathrm{E}[Y_t|Y_0] = Y_0$ and variance $\mathrm{Var}(Y_t|Y_0) = t\sigma_\varepsilon^2$.

From the point of view of local asymptotic experiments, however, Models (a) and (b) differ dramatically. While Model (a), as we shall see, defines local experiments that are nicely (be it with nonstandard $n^{3/2}$ consistency rates) LAN (Locally Asymptotically Normal) at the null hypothesis H_0 of unit root[3], Model (b) at $H_0^{(b)}$ yields a considerably more tricky asymptotic structure, of the LABF (Locally Asymptotically Brownian Functional) type, for which no uniform optimality results exist—see Elliot, Rothenberg, and Stock (1996), Rothenberg and Stock (1997), Thompson (2004), and Jansson (2008). We refer to Gushchin (1996), Ploberger (2004, 2008), and Jansson and Moreira (2006), for recent developments on experiments of the LABF and the (more general) LAQ (Locally Asymptotically Quadratic) type.

For any fixed n, thus, the differences between Model (a) and (b) are extremely tenuous: for $\rho < 1$, they strictly coincide, whereas, for $\rho = 1$, Model (a) is more general, since H_0 includes $H_0^{(b)}$ as a special case. It follows that the choice between (1) and (2) is not really a choice between two models, but a choice between two types of asymptotics: the debate is about (a)-asymptotics versus (b)-asymptotics rather than Model (a) versus Model (b). This is a debate we do not enter into here. Asymptotics in this paper are just a mathematical device, which is used to suggest "sensible" testing procedures for the finite-sample problem at hand. Rather than parametric or semiparametric efficiency, or ARE values, which presuppose a specific asymptotic scheme, the ultimate benchmark for the procedures we are describing here are their finite-sample performance under the alternative, where Models (a) and (b) coincide, so that no particular choice needs to be made.

1.2. Outline of the paper

The remainder of the paper accordingly is organized in two main parts: Section 2, which is devoted to asymptotics, and Section 3, dealing with finite-sample performances.

Much attention has been given, in the recent literature, to (b)-asymptotics. The analysis we are developing in Section 2 is based on (a)-asymptotics[4], which, apparently, have

[3]with degenerate Fisher information at $\mu = 0$, though.

[4]We once more emphasize that asymptotics here are just an agnostic mathematical device, the consequences of which are to be evaluated (Section 3) on the basis of finite-sample performances.

not been considered so far in this context, and suggest a class of very simple tests, for which moreover rank-based, hence finite-sample distribution-free versions, exist. Being distribution-free, those tests are valid, for finite sample size n, irrespective of the innovation density f (no moment restrictions[5]), and irrespective of the model ((a) or (b)). We provide a full analysis of the limiting properties of those tests: asymptotic null distributions and, under (a)-asymptotics, local powers and asymptotic relative efficiencies (AREs).

Section 3 is devoted to a numerical investigation of the finite-sample performance of the tests described in Section 2—an investigation that does not require any choice between Model (a)- or (b), as both models describe the same data-generating processes under the alternative. That finite-sample analysis brings into the picture an important new feature of the problem: the influence of the initial observation Y_0. Müller and Elliott (2003) show that the deviation of Y_0 from the stationary mean has a dramatic influence on the finite-sample performance of all unit-root tests. In empirical applications it is generally impossible to tell whether that deviation is small or large. Elliott and Müller (2006) provide a discussion for this; in Section 3.2 below, we are following their suggestion of evaluating empirical performances as a function of $Y_0 - m$ by adopting their simulation design. The results show that our rank tests significantly outperform all their competitors (the traditional Dickey-Fuller procedures, as well as the tests by Elliot, Rothenberg, and Stock (1996), Ng and Perron (2001), and Elliott and Müller (2006)) whenever the deviation $Y_0 - m$ of the initial value Y_0 from the stationary mean is "large", and whenever the innovation distribution is heavy-tailed.

Section 4 concludes, while proofs are gathered in an Appendix.

1.3. Rank tests

Before turning to asymptotics, let us provide some details about the rank-based tests we are proposing. Our test statistics are based on the ranks R_t of the increments $\Delta Y_t := Y_t - Y_{t-1}$. Let g be a given density (the so-called *reference density*), not necessarily the actual underlying one f. We assume throughout that g belongs to the class $\mathcal{F}$ of densities h that are absolutely continuous with a.e. continuous derivative h' and finite

[5]In the absence of first-order moments, m and μ can be reinterpreted as medians rather than means.

Fisher information for location $I_h := \int (h'/h)^2 \, \mathrm{d}H \in (0, \infty)$, and for which

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{h'}{h} \left(H^{-1}\Big(\frac{i}{n+1}\Big) \right) \right\}^2 = I_h \tag{7}$$

(as usual, F, G, H denote the distribution functions associated with f, g, h).

We stress again that, as far as the validity of our test is concerned, we do not make *any* assumptions on f (our tests are strictly distribution-free). If, however, asymptotic optimality, under density f and (a)-asymptotics, is to be considered, then we need to impose $f \in \mathcal{F}$.

Motivated by the asymptotic analysis of Section 2, our test statistics take the form

$$T_g^{(n)} := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \varphi_g \left(\frac{R_t}{n+1} \right), \tag{8}$$

with $\varphi_g(u) := -g'\left(G^{-1}(u)\right)/g\left(G^{-1}(u)\right)$, $u \in (0,1)$. Under the null hypothesis H_0, hence also under the null hypothesis $H_0^{(b)}$, the vector of ranks $(R_1, \ldots, R_n)$, and therefore the test statistics $T_g^{(n)}$, are distribution-free with respect to μ and f. In particular, this implies that exact critical values for $T_g^{(n)}$-based tests can be easily computed or simulated for finite n, despite the unspecified f and μ.

The form of the test statistic (8) actually follows from optimality considerations under (a)-asymptotics and $\mu \neq 0$. In Section 2, we derive its local power and compare it to the efficiency bound obtained from the LAN property (derived in Section 2.3). That local power does depend on both the reference density g and the actual underlying density f. We show that a correctly specified reference density $g = f$ leads to a test that achieves the efficiency bound and thus is *parametrically* efficient. As a result, while our tests are valid irrespective of the reference and underlying densities, they are locally and asymptotically efficient, in Model (a) (with $\mu \neq 0$), in case of a correctly specified g.

This situation thus is tantamount to quasi- or pseudo-maximum likelihood estimation, where choosing a (Gaussian) reference density leads to an estimator that (often) remains consistent even when the reference density is misspecified, while attaining the parametric efficiency bound in case the actual underlying density is Gaussian. In general, the limiting variance of such estimators, however, depends on both the true and the (Gaussian) reference density. Our tests have a comparable property, with the important difference that we may use *any* density g as a reference density, while quasi or pseudo likelihood procedures are generally restricted to a Gaussian g (when using another reference density the estimators, in general, do not remain consistent under misspecified innovation density). Moreover, for our tests, the reference density can even be pre-estimated in order to achieve (parametric) efficiency uniformly over a broad class of densities f—without any sacrifice at the level of validity (see Section 2.6).

Now, in case (b)-asympotics are to be preferred, the tests based on $T_g^{(n)}$, as already mentioned, remain valid; but their asymptotic optimality properties are lost. However, their fixed-alternative performances are unchanged: see Section 3.

Distribution-freeness is another attractive property of our tests. The need for exact and distribution-free inference in econometrics often has been emphasized: see, for instance, Dufour (1997) or Coudin and Dufour (2009). Despite of that recognized need, distribution-free procedures remain extremely rare in the context of time series econometrics. Campbell and Dufour (1995), Campbell and Dufour (1997), and Luger (2003) consider testing orthogonality restrictions using sign- and rank-based tests instead of regression-based approaches. These methods are based on zero-median or symmetry assumptions and, using extensive simulation, are shown to beat regression-based tests. Hasan and Koenker (1997) extend these results using regression rank-scores in order to deal with the nuisance parameter problem. Their focus of interest again is the zero-mean unit root model. Hasan (2001) further allows for infinite variances; no formal optimality analysis is given. Thompson (2004b) reconsiders these tests in order to improve their power, especially under fat-tailed error distributions. Finally, we mention Breitung and Gouriéroux (1997) who consider the hypothesis that some transformation of the process exhibits a unit root. They propose a test based on the ranks of the observed time series (not those of residuals).

2. Asymptotic theory

2.1. Rank tests: exact versus approximate scores

It turns out that deriving results on the asymptotic size and (under (a)-asymptotics) local power of our test is easier when the test statistic (8) is slightly adjusted, replacing φ_g by

$$\tilde{\varphi}_g(u) := \mathrm{E}_G \left\{ \varphi_g \left(G \left(\varepsilon_t \right) \right) \middle| R_t = \lfloor u(n+1) \rfloor \right\}, \quad u \in (0,1). \tag{9}$$

Note that $\tilde{\varphi}_g$, contrary to φ_g, depends on the number of observations n. Clearly, the statistic based on φ_g is simpler to compute, although the function $\tilde{\varphi}_g$ is easily simulated using distribution-freeness of the ranks. Whereas (8), in the literature on rank-based inference, is known as the *approximate score* version of $T_g^{(n)}$, using $\tilde{\varphi}_g$ in $T_g^{(n)}$ yields the so-called *exact score* version. This exact score version is more convenient for proofs as its expectation is identically zero irrespective of the true underlying density f: $\mathrm{E}\{\tilde{\varphi}_g(R_t/(n+1))\} = \mathrm{E}_G\{\varphi_g(G(\varepsilon_t))\} = 0$. Incidentally, note that the average of the weighting constants $t/(n+1) - 1/2$ in (8) equals zero as well. When n is large and conditionally on the rank of ε_t being $R_t = i$, $G(\varepsilon_t)$ is approximately equal to $i/(n+1)$. This intuitively explains why the φ_g- and $\tilde{\varphi}_g$-based versions of $T_g^{(n)}$ behave similarly. This is formalized in the following result.

Lemma 2.1 *If the reference density g belongs to $\mathcal{F}$, we have, as $n \to \infty$, under the null hypothesis H_0 of unit root,*

$$T_g^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \tilde{\varphi}_g \left(\frac{R_t}{n+1} \right) + o_{\mathrm{P}}(1). \tag{10}$$

Proof: This is a well-known result on the asymptotic equivalence of the *approximate* and *exact* score versions of (linear) rank statistics, which is proved at various places; see, for instance, Theorem 13.5 in Van der Vaart (2000). □

Remark 2.1 *A consequence of the Local Asymptotic Normality result proved in Proposition 2.1 below is mutual contiguity of the probability measures at the unit root ($\rho = 1$) and those near the unit root ($\rho_n = 1 - O(n^{-3/2})$). The asymptotic equivalence (10), therefore, is preserved under contiguous sequences. Consequently, in expressions like (10), we do not have to worry whether o_{P}'s are taken at the unit root or near the unit root. This consequence of contiguity will be used throughout the paper without further mention.*

Condition (7) on φ_g is satisfied for all standard reference densities g: Gaussian, logistic, double-exponential, Student (including Cauchy), etc. Under this condition, the asymptotic equivalence in (10) implies that all results concerning asymptotic size, power (under contiguous alternatives), and efficiency carry over from one statistic to the other: whether exact or approximate scores are considered has no impact on asymptotic results.

2.2. Rank tests: Asymptotic size

In view of distribution-freeness, one easily constructs, via simulations, tests based on $T_g^{(n)}$ with exact finite-sample sizes, irrespective of μ and f. Asymptotic critical values can be obtained from a normal distribution with variance $I_g/12$, as shown by the following result (see the appendix for a proof).

Theorem 2.1 *Let* $(\varepsilon_1, \dots, \varepsilon_n)$ *be i.i.d. from a continuous distribution with density* f *and denote by* R_t *the rank of* ΔY_t *among* $\Delta Y_1, \dots, \Delta Y_n$*. Let the reference density* g *belong to* $\mathcal{F}$*. Then, as* $n \to \infty$ *and under* H_0*,*

$$\sqrt{12/I_g} T_g^{(n)} \Rightarrow \mathcal{N}(0,1). \tag{11}$$

Note that $\sqrt{12/I_g} T_g^{(n)}$ is scale-free. If σ is a scale parameter associated with g (not necessarily a standard error, though), writing g_σ for g and g_1 for the corresponding standardized density (such that $g_\sigma(x) = \frac{1}{\sigma} g_1(\frac{x}{\sigma})$), we have indeed $\sqrt{12/I_{g_\sigma}} T_{g_\sigma}^{(n)} = \sqrt{12/I_{g_1}} T_{g_1}^{(n)}$.

We insist, once again, that no assumptions are made on f which, in particular, needs not have finite moments nor belong to $\mathcal{F}$. Moreover, Theorem 2.1 is equally valid for Model (a) as well as Model (b) as a result of the distribution-freeness also with respect to μ. For instance, Theorem 2.1 still applies under heavy-tailed innovations such as Cauchy or Lévy ones, while the Dickey-Fuller statistic may break down. This fact will be confirmed in Section 3 by finite-sample simulations. Unlike their size, however, the power of our tests depends both on the chosen reference density g and the actual underlying density f (actually, on their standardized versions, g_1 and f_1); for $f \in \mathcal{F}$, explicit values are provided in Theorem 2.2 below.

2.3. Limit experiment and efficient inference

As mentioned in the introduction, the limiting experiments, under (a)-asymptotics, crucially depend on the value of μ, leading to (a)-asymptotics for $\mu \neq 0$ and to (b)-asymptotics for $\mu = 0$. In the latter case, the limit experiment (for the model with single parameter ρ) is Locally Asymptotically Brownian Functional (LABF) with rate of convergence n, as shown by Jeganathan (1995), and departures of the order of $n^{-3/2}$ from the unit-root hypothesis cannot be detected. This LABF-result is exploited in Jansson (2008) to derive power envelopes for unit root tests.

As shown in the next result, the situation is quite different, and much simpler, under (a)-asymptotics at rate $n^{-3/2}$.

Proposition 2.1 *Consider Model (a) with innovation density* $f \in \mathcal{F}$*, and denote by* $\mathrm{P}^{(n)}_{(\mu,\rho);f}$ *the joint distribution of* $(Y_1, \dots, Y_n)$ *under (1).*

(i) The family $\{\mathrm{P}^{(n)}_{(\mu,\rho);f} \,|\, \mu \in \mathbb{R},\ \rho \in [-1,1]\}$ *is Locally Asymptotically Normal (LAN) at any* $(\mu, \rho = 1)$*, for local alternatives of the form* $(\mu_n = \mu + h_1 n^{-1/2}, \rho_n = 1 + h_2 n^{-3/2})$*,*

with central sequence

$$\left[\begin{array}{c} \Delta_\mu^{(n)} \\ \Delta_\rho^{(n)} \end{array}\right] := \left[\begin{array}{c} n^{-1/2}\sum_{t=1}^{n} \frac{-f'}{f}(\Delta Y_t) \\ \mu n^{-1/2}\sum_{t=1}^{n} \frac{t}{n+1}\frac{-f'}{f}(\Delta Y_t) \end{array}\right] \tag{12}$$

and Fisher information

$$I_f \left[\begin{array}{cc} 1 & \mu/2 \\ \mu/2 & \mu^2/3 \end{array}\right]. \tag{13}$$

More precisely, $\Delta Y_t = \mu + \varepsilon_t$ *under* $\mathrm{P}^{(n)}_{(\mu,1);f}$*, and, as* $n \to \infty$*,*

$$\begin{aligned} \log \frac{d\mathrm{P}^{(n)}_{(\mu_n,\rho_n);f}}{d\mathrm{P}^{(n)}_{(\mu,1);f}} &= h_1 n^{-1/2} \sum_{t=1}^{n} \frac{-f'}{f}(\varepsilon_t) + h_2 \mu n^{-1/2} \sum_{t=1}^{n} \frac{t}{n+1}\frac{-f'}{f}(\varepsilon_t) \\ &\quad -\frac{I_f}{2}\left(h_1^2 + \mu h_1 h_2 + \frac{\mu^2}{3}h_2^2\right) + o_{\mathrm{P}}(1) \end{aligned}$$

and

$$\left[\begin{array}{c} \Delta_\mu^{(n)} \\ \Delta_\rho^{(n)} \end{array}\right] \Rightarrow N\left(0, I_f \left[\begin{array}{cc} 1 & \mu/2 \\ \mu/2 & \mu^2/3 \end{array}\right]\right).$$

For $\mu = 0$*, however, this LAN result is a degenerate one, with information matrix* $I_f \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array}\right]$.

(ii) If f *has finite variance, the subfamily* $\left\{\mathrm{P}^{(n)}_{(\mu,\rho);f} \mid \mu = 0,\ \rho \in [-1,1]\right\}$ *is Locally Asymptotically Brownian Functional (LABF) for local alternatives of the form* $\rho_n = 1 + h_2 n^{-1}$.

Proof: See the Appendix. □

Remark 2.2 *The LAN result of Proposition 2.1 does not require* $h_2 \leq 0$*: all claims in this paper can easily be rephrased in the context of testing* $H_0:\ \rho = 1$ *against* $H_1:\ \rho > 1$ *and* $H_0:\ \rho = 1$ *against* $H_1:\ \rho \neq 1$.

Remark 2.3 *In case one considers the model* $Y_t = \rho Y_{t-1} + \mu + \gamma t + \varepsilon_t$*, i.e. a model including a linear time-trend, the LAN result still holds true when* $\gamma \neq 0$*, but with consistency rate (for* ρ*)* $n^{5/2}$ *instead of* $n^{3/2}$.

Remark 2.4 *The fact that the Fisher information for* ρ *in (13) vanishes for* $\mu \to 0$ *confirms that* ρ *indeed cannot be estimated at rate* $n^{3/2}$ *whenever* $\mu = 0$.

Remark 2.5 *An initial value* Y_0 *with distribution depending on* ρ*, such as*

$$Y_0 \sim \mathcal{N}(\mu/(1-\rho), \sigma_f^2/(1-\rho^2)),$$

can deteriorate the LAN result. In such situations, our LAN result still holds conditional on Y_0. In this way one ignores the statistical information possibly contained in Y_0, and restricts attention to the differenced observations $\Delta Y_1, \ldots, \Delta Y_n$.

Local Asymptotic Normality, via the Hájek and Le Cam asymptotic theory of statistical experiments (see, e.g., Chapters 7 and 9 of Van der Vaart (2000)) completely characterizes the local and asymptotic features of the statistical experiment under study. Not only does it induce the asymptotic optimality bounds for statistical inference, but it also indicates how central-sequence-based procedures achieve those bounds. Accordingly, it follows from Proposition 2.1 that a locally and asymptotically optimal test for $H_0 : \ \rho = 1$, under (a)-asymptotics, in case the innovation density f is known, and considering $\mu \neq 0$ a nuisance parameter, should be based on (any monotone transformation of)

$$I_f^{-1}\left(\Delta_\rho^{(n)} - \frac{\mu}{2}\Delta_\mu^{(n)}\right) = \frac{\mu}{I_f} n^{-1/2} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2}\right) \frac{-f'}{f}\left(\Delta Y_t\right) \qquad (14)$$

(see, for instance, Section 11.9 of Le Cam (1986)). Clearly, the magnitude of the constant factor μ/I_f can be ignored in the construction of that test. Since the sign of μ is unspecified, both one- and two-sided versions are meaningful. In the remainder of this section, we focus on the empirically more relevant case of $\mu > 0$; asymptotic theory then leads to rejecting (as the alternative is $\rho < 1$) for small values of the test statistic. In Section 3, however, we evaluate finite-sample performance for $\mu = 0$, and consider two-sided tests. Statistics of the form

$$S_g^{(n)} := n^{-1/2} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2}\right) \frac{-g'}{g}\left(\Delta Y_t\right) \qquad (15)$$

thus are interesting candidates as test statistics for our problem, and reach parametric efficiency in case $f = g$. Unfortunately, $S_g^{(n)}$ is not distribution-free.

The situation is totally different if we turn to $T_g^{(n)}$. Under $f = g$, indeed, it follows from (15), (23) and Lemma A.1 that $T_g^{(n)} = S_g^{(n)} + o_{\mathrm{P}}(1)$ under H_0 and $f = g$. In case the actual density coincides with g, $T_g^{(n)}$ thus shares all the nice optimality features of $S_g^{(n)}$. The essential difference is that, being distribution-free, its finite-sample null distribution is the same under $f \neq g$ as under $f = g$: $T_g^{(n)}$ thus does not require f to be specified, and naturally qualifies as a solution for our testing problem, while achieving efficiency at the chosen reference density g.

2.4. Local powers

The asymptotic power of our rank-based test statistics $T_g^{(n)}$ against local (under (a)-asymptotics) unit root alternatives follows directly from the so-called Le Cam third lemma, provided that f and g both satisfiy the assumptions of Proposition 2.1.

Theorem 2.2 *Consider the model (1) with innovation density $f \in \mathcal{F}$ and $Y_0 \sim \mathcal{L}$. Let the reference density g also be in $\mathcal{F}$. Then, under $\mathrm{P}^{(n)}_{(\mu,\rho_n);f}$, where $\rho_n = 1 + hn^{-3/2}$,*

$$T_g^{(n)} \Rightarrow \mathcal{N}(h\mu I_{fg}/12, I_g/12) \quad \text{as } n \to \infty, \tag{16}$$

with

$$I_{fg} := \int_{u=0}^{1} \varphi_g(u)\varphi_f(u)du. \tag{17}$$

Proof: See the Appendix. □

Whenever $\mu \neq 0$, our test has power against alternatives that are at distance $n^{-3/2}$ from the unit root. This is, of course, much more precise than the usual $n^{-1/2}$ rate. It is more precise, too, than the n^{-1} rate that can be attained in case $\mu = 0$, see Proposition 2.1. In that case, however, no test can have local power against alternatives at rate $n^{-3/2}$.

It is interesting to compare (still under (a)-asymptotics) the power of our test statistic to that of the classical Dickey-Fuller test. For this comparison we choose the asymptotically optimal Dickey-Fuller test for Model (a), that is, based on the least-squares estimate $\hat{\rho}_n^{DF}$ of ρ in (1). The asymptotic properties of this classical Dickey-Fuller statistic are well known and we have the following corollary to Theorem 2.2.

Corollary 2.1 *Let f and g belong to $\mathcal{F}$; assume $\mu > 0$ and that f moreover has finite variance σ_f^2. The Asymptotic Relative Efficiency, for the unit root hypothesis $H_0:\ \rho = 1$, of the one-sided rank test based on $T_g^{(n)}$ with respect to the Dickey-Fuller test based on $\hat{\rho}_n^{DF}$ is, under density f,*

$$ARE_f(T_g^{(n)}|DF) = |I_{fg}|^3\sigma_f^3/I_g^{3/2}. \tag{18}$$

Proof: See the Appendix □

Remark 2.6 *The ARE_f in (18) is defined as the limit, as $n \to \infty$, of the ratio n_{DF}/n, where n_{DF} is the number of observations needed in the Dickey-Fuller test to achieve the same performance (in terms of power) as of our rank-based test using n observations.*

	Actual density f				
Reference density g	Gaussian	logistic	DExp	t_3	Cauchy
Gaussian (van der Waerden)	1.00	1.07	1.44	2.10	∞
logistic (Wilcoxon)	0.93	1.15	1.84	2.62	∞
double exponential (Laplace)	0.51	0.75	2.83	2.06	∞

Table 1: Asymptotic Relative Efficiencies $\mathrm{ARE}_f(T_g^{(n)}|\mathrm{DF})$ of our rank-based test based on $T_g^{(n)}$ in (8) with respect to the Dickey-Fuller test, for various choices (Gaussian, logistic, double exponential) of the reference density g, and several values (Gaussian, logistic, double exponential, Cauchy, and t_3) of the actual density f.

Our test and the Dickey-Fuller test both have local power at rate $n^{3/2}$. This explains the exponent three in (18).

Remark 2.7 *Despite the notation, ARE_f in (18) is a scale-free quantity. It is easy to see, indeed, that, writing f_1 for the standardized version of f (that is, $f(z) = \sigma_f^{-1} f_1(z/\sigma_f)$), $I_{fg}^3 \sigma_f^3 = I_{f_1 g}^3$. Similarly, if g_1 and g_2 are such that, for some $c > 0$, $g_2(z) = c^{-1} g_1(z/c)$, then $I_{fg_2}/I_{g_2}^{1/2} = I_{fg_1}/I_{g_1}^{1/2}$.*

Table 1 provides, for various reference densities and various f, some numerical values of (18). Under infinite innovation variance, those values are infinite, since Dickey-Fuller is no longer valid[6]. Inspection of Table 1 reveals that, under finite innovation variance for f, very sizeable efficiency gains also are possible, even when using a Gaussian reference density g (van der Waerden tests).

2.5. Choosing a reference density g

Our test depends on a reference density g to be chosen by the investigator. This raises the obvious question of how to choose this reference density.

Recall that our rank-based statistic $T_g^{(n)}$ is homogeneous in the scale of the reference distribution: rescaling a given reference density $g(\cdot)$ to $g_c(\cdot) = c^{-1} g(\cdot/c)$, $c > 0$ has no impact on the test, and one does not have to worry about choosing an appropriate scale for g. Similarly, we have shown in Remark 2.7 that the Asymptotic Relative Efficiency of our test with respect to the Dickey-Fuller test does not depend on the scale of the reference density g, nor on that of the actual density f.

[6] Recall that for symmetric α-stable innovation distributions the Dickey-Fuller test statistic has a limiting distribution of the Lévy-type with critical values dependent on the tail index α; see Rachev, Mittnik, and Kim (1998), Ahn, Fotopoulos, and He (2003), and Callegari, Cappuccio, and Lubian (2003).

The form of the reference density g, if not its scale, however, does influence the local power of our test via the ratio $|I_{fg}|/I_g^{1/2}$ in (17). An obvious first choice is a Gaussian reference density $g(x) \propto \exp(-x^2/2)$, leading to the so-called *normal* or *van der Waerden* scores. In this case,

$$T_{\mathrm{vdW}}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \Phi^{-1} \left(\frac{R_t}{n+1} \right),$$

where Φ denotes the standard normal distribution function, $I_g = 1$, and (18) reduces to

$$\mathrm{ARE}_f(\mathrm{vdW}|\mathrm{DF}) = \left| \int_{u=0}^{1} \frac{-f'}{f} \left(F^{-1}(u) \right) \Phi^{-1}(u) \right|^3 \sigma_f^3. \tag{19}$$

A celebrated result by Chernoff and Savage (1958) shows that the latter quantity is always larger than one, except under Gaussian f, where it takes value one. Consequently, a Gaussian reference density constitutes a safe choice, as it always leads to an improvement over the Dickey-Fuller test. The magnitude of the improvement is all the more sizeable in our situation, due to the faster rate of convergence $n^{3/2}$; see the first row in Table 1. For instance, true underlying Student t_3 distributed innovations lead to more than 100% efficiency gain, while fatter-than-t_3-tailed distribution lead to even larger (infinite in the case of infinite innovation variance) gains.

Two other popular choices for the reference density are the Double Exponential distribution (Laplace or sign test scores), with density $g_L(x) = \exp(-\sqrt{2}|x|)/\sqrt{2}$ (for which $\sigma_{g_L}^2 = 1$ and $I_{g_L} = 2$), and the logistic distribution (Wilcoxon scores) $g_W(x) = \pi \exp(-\pi x/\sqrt{3})/(\sqrt{3}(1 + \exp(-\pi x/\sqrt{3}))^2)$ (for which $\sigma_{g_W}^2 = 1$ and $I_{g_W} = \pi^2/9$). They lead to the Laplace and Wilcoxon test statistics

$$\begin{aligned} T_L^{(n)} &= \sqrt{\frac{2}{n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \mathrm{sgn} \left(\frac{R_t}{n+1} - \frac{1}{2} \right), \\ T_W^{(n)} &= \frac{\pi}{\sqrt{3n}} \sum_{t=1}^{n} \left(\frac{t}{n+1} - \frac{1}{2} \right) \frac{1 - \frac{n+1-R_t}{R_t}}{1 + \frac{n+1-R_t}{R_t}}, \end{aligned}$$

respectively.

It is worth emphasizing, again, that we nowhere impose that the innovations need to have finite variances, nor even finite first-order moments: our tests remain valid under completely unspecified innovation density f and completely unspecified shift μ (which may be zero). As explained before, the Dickey-Fuller test is no longer valid in the semi-parametric model with unspecified f.

Remark 2.8 *In view of Theorem 2.2, for given f, maximum power is achieved when the reference density g matches the actual one f (up to a possible scale transformation). In that case, our rank-based statistic asymptotically coincides with the parametrically optimal (under (a)-asymptotics) test statistic (14), and the $T_g^{(n)}$-based test achieves parametric efficiency in Model (a) with innovation density f. This implies that Model (a) (with innovation density f) actually is* adaptive*: the "cost" of not knowing the innovation density in addition to not knowing μ is asymptotically nil when performing inference about ρ. Model (b) does not exhibit such attractive limiting local structure.*

2.6. Pre-estimating the reference density g

As the power of the test depends on the chosen reference density, and is maximal if the reference density coincides with the actual density f up to a scale transformation, one may want to pre-estimate the reference density to use. An important additional advantage of our test is that this can be done without any changes in the asymptotic analysis.

To be more precise, consider an estimated reference density $\hat{g}_n$ with values in $\mathcal{F}$ that depends on the order statistics of the increments ΔY_t, as is, for example, the case for traditional kernel density estimators. Recall that the order statistics are stochastically independent of the ranks R_t of the innovations. Therefore, we can easily study the behavior of $T_{\hat{g}_n}^{(n)}$ conditionally on the order statistics, that is, as if $\hat{g}_n \in \mathcal{F}$ were a given reference density. In particular, if (conditionally on the order statistics) exact α-critical points are computed for the estimated-score version of (8), conditional size, hence also the unconditional one, is exactly α too. The resulting tests moreover have Neyman α-structure with respect to the order statistics, hence are similar and unbiased. An analogous reasoning can be applied to show that the power properties of our test with estimated reference density are as if the reference density were correctly specified. In order to make sure that $I_{\hat{g}_n}$ converges to I_g a construction as in Proposition 7.8.1 in Bickel, Klaassen, Ritov, and Wellner (1993) can be considered.

Summing up, the tests based on $T_{\hat{g}_n}^{(n)}$ remain conditionally distribution-free; they are parametrically efficient (under (a)-asymptotics), uniformly over the family of all $\mu \neq 0$ and all f such that, under f, $T_{\hat{g}_n}^{(n)} - T_f^{(n)} = o_{\mathrm{P}}(1)$ —without losing finite-sample validity over that broader class of *all* μ and f.

3. Finite-sample performance

As mentioned in the introduction, the ultimate benchmark for any statistical procedure is its finite-sample performance. This is all the more true in the present context, where two distinct and plausible asymptotic schemes are coexisting, roughly on the same statistical model. This section is totally agnostic in that respect, and does not make any choice between (a)- and (b)-asymptotics. Nevertheless, the description of the simulated data-generating process requires a parameterization, and, without any loss of generality, the $(\rho,\ m)$ parameterization (2) is adopted throughout.

Section 3.1 deals with the finite-sample behavior of our tests under H_0, hence, a fortiori, also under $H_0^{(b)}$. Section 3.2 discusses their behavior under alternatives (where Model (a) and Model (b) coincide).

3.1. Finite-sample sizes

It follows from Theorem 2.1 that the rank-based test statistic $T_g^{(n)}$ is asymptotically $\mathcal{N}(0, I_g/12)$ under the null hypothesis. This section studies the finite-sample null distribution of $T_g^{(n)}$. Recall once more that our rank-based test statistics are distribution-free under the null hypothesis. This means that the finite-sample distribution of $T_g^{(n)}$ only depends on the number n of observations and the choice of the reference density g. Such distributions can easily be tabulated.

To illustrate the convergence to a $\mathcal{N}(0, I_g/12)$ distribution under the null hypothesis, Figure 1 presents a scaled histogram of simulated values of $T_{\mathrm{vdW}}^{(n)}$ along with its limiting Gaussian density for $n = 25,\ 50,\ 100$. From the figure we conclude that the convergence to

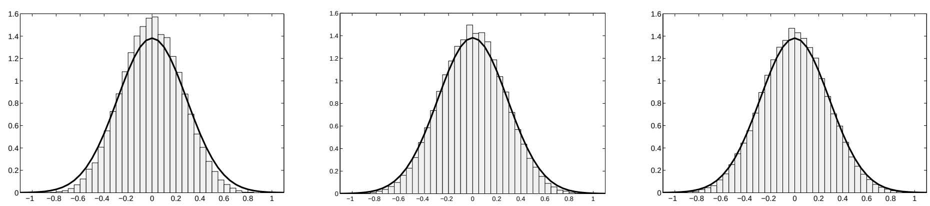

Figure 1: Simulated (50,000 replications) finite-sample ($n = 25,\ 50,\ 100$) distributions of the van der Waerden test statistic $T_{\mathrm{vdW}}^{(n)}$ (reference density $g = \phi$), compared to its limiting distribution under the null hypothesis.

the limiting distribution is quite fast. This is common for rank-based statistics. Moreover, in view of distribution-freeness, this convergence is uniform over the family of possible underlying innovation densities f, irrespective of μ. Note that the limiting distribution seems to be overestimating tail probabilities, hence produces conservative critical values. This is confirmed by Table 2, where simulated quantiles are presented for various sample sizes n and various reference densities g, along with (in the rows labeled "$n = \infty$") the asymptotic ones. As the distributions are symmetric with respect to the origin, only right-tail quantiles are presented.

Although the convergence is fast, we thus recommend using simulated critical values rather than the asymptotic ones.

Reference density g		Gaussian (van der Waerden)	logistic (Wilcoxon)	double exponential (Laplace)
	$n = 25$	0.62	0.71	0.99
	$n = 50$	0.68	0.75	1.02
$q = 0.5\%$	$n = 100$	0.70	0.76	1.04
	$n = 250$	0.73	0.77	1.04
	$n = \infty$	0.74	0.78	1.05
	$n = 25$	0.49	0.56	0.76
	$n = 50$	0.52	0.57	0.78
$q = 2.5\%$	$n = 100$	0.54	0.58	0.79
	$n = 250$	0.55	0.59	0.80
	$n = \infty$	0.57	0.59	0.80
	$n = 25$	0.41	0.47	0.65
	$n = 50$	0.44	0.48	0.66
$q = 5\%$	$n = 100$	0.45	0.49	0.67
	$n = 250$	0.46	0.49	0.67
	$n = \infty$	0.47	0.50	0.67

Table 2: Simulated $(1 - q)$-quantiles (based on 50,000 replications) for the van der Waerden, Wilcoxon, and Laplace rank-based test statistics, various values of n and q, under H_0, hence, a fortiori, also under $H_0^{(b)}$. The rows labeled "$n = \infty$" contain the critical values calculated from the limiting Gaussian distribution.

3.2. Finite-sample powers

As discussed in the introduction, the ultimate benchmark for any statistical procedure is its finite-sample performance. This is all the more true in the present context, where several distinct and plausible asymptotic schemes are coexisting, roughly on the same statistical model. This section is totally agnostic in that respect, and does not make any choice between Models (a) and (b), nor between the corresponding asymptotics. Nevertheless, the description of the simulated data-generating process requires a parameterization, and the $(\rho,\ m)$ parameterization (2) is adopted throughout. As mentioned in the introduction, the initial value Y_0 or, more precisely, its deviation $Y_0 - m$ from the stationary mean (a quantity which, in practice, is not known), heavily influences the power of all unit-root tests. Following Elliott and Müller (2006) we therefore explore powers for various values of Y_0, of the form $Y_0 = m + a\sigma_\varepsilon/\sqrt{1-\rho^2}$ with $a = 0, 1, \ldots, 6$ $(\rho < 1)$ measuring the amplitude of the deviation of Y_0 from the stationary mean in terms of the stationary standard deviation[7].

Tables 3-10 below provide rejection frequencies, over 25,000 replications of the data-generating process, and sample sizes $n = 50$ and $n = 100$, of three of the rank-based tests (van der Waerden, Wilcoxon, and Laplace, associated with Gaussian, logistic and double-exponential reference density g, respectively) considered in this paper, along with those of the traditional Dickey-Fuller procedure, the P_T-test $(c = -7)$ ERS-P_T from Elliot, Rothenberg, and Stock (1996), the M^{GLS} tests NP-MZ_α^{GLS}, NP-MZ_t^{GLS}, and NP-MSB^{GLS} (with $p = 0$ and $\bar{c} = -7.0$), from Ng and Perron (2001), and the $\hat{Q}^\mu$ tests EM-$\hat{Q}^\mu(10, 1)$ and EM-$\hat{Q}^\mu(10, 3.8)$ from Elliott and Müller (2006). Throughout, the nominal level is $\alpha = 5\%$, with simulated critical values for the rank-based tests and asymptotic critical values for the other ones. As all tests are invariant with respect to m (under the null as well as under the alternative), we only consider $m = 0$. For each combination of an innovation density f (four densities: Gaussian, double-exponential, Cauchy, and skew-normal) and a ρ value (four values: 1, 0.99, 0.975, and 0.95), following Elliott and Müller (2006), seven starting values $(Y_0 = a\sigma_\varepsilon/\sqrt{1-\rho^2}$ for $a = 0, 1, \ldots, 6)$ have been considered.[8] All simulations were carried out in Matlab 7.10; codes are available upon request.

In each table, rejection frequencies significantly larger than 10% (at probability level $\alpha = 5\%$, that is, larger than or equal to 0.097) are printed in boldface; among them, the winners in each column (still at level $\alpha = 5\%$) are starred.

Before commenting the results, some further details about the implementation of Dickey-Fuller are in order. The Dickey-Fuller tests actually are the (standard) t-tests

[7]For $\rho = 1$, that deviation is not well-defined; all test statistics, however, only depend on the observations via $\Delta Y_1, \ldots, \Delta Y_n$ which, under the null, coincide with $\varepsilon_1, \ldots, \varepsilon_n$, so that, without any loss of generality, we put $Y_0 = 0$, in simulations under the null.

[8]For the Cauchy density, we use $\sigma_\varepsilon = 3$ in the definition of Y_0.

Table 3: $n = 50$; $25,000$ replications; $f = \mathcal{N}(0,1)$

Test	a						
	0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.081	0.081	0.081	0.081	0.081	0.081	0.081
ERS-P_T	0.048	0.048	0.048	0.048	0.048	0.048	0.048
NP-MZ_α^{GLS}	0.058	0.058	0.058	0.058	0.058	0.058	0.058
NP-MSB_α^{GLS}	0.053	0.053	0.053	0.053	0.053	0.053	0.053
NP-MZ_t^{GLS}	0.052	0.052	0.052	0.052	0.052	0.052	0.052
EM-$\hat{Q}^\mu(10,1)$	0.025	0.025	0.025	0.025	0.025	0.025	0.025
EM-$\hat{Q}^\mu(10,3.8)$	0.026	0.026	0.026	0.026	0.026	0.026	0.026
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.081	0.080	0.081	0.080	0.080	0.081	0.080
ERS-P_T	0.060	0.054	0.038	0.020	0.009	0.003	0.001
NP-MZ_α^{GLS}	0.072	0.066	0.044	0.025	0.010	0.004	0.001
NP-MSB_α^{GLS}	0.065	0.060	0.042	0.024	0.011	0.004	0.002
NP-MZ_t^{GLS}	0.067	0.060	0.040	0.022	0.009	0.003	0.001
EM-$\hat{Q}^\mu(10,1)$	0.031	0.029	0.021	0.011	0.006	0.002	0.001
EM-$\hat{Q}^\mu(10,3.8)$	0.029	0.032	0.026	0.018	0.013	0.007	0.003
van der Waerden	0.047	0.047	0.048	0.049	0.052	0.055	0.058
Laplace	0.049	0.050	0.050	0.050	0.053	0.055	0.058
Wilcoxon	0.047	0.048	0.049	0.051	0.052	0.055	0.060
$\rho = 0.975$							
Dickey-Fuller	0.084	0.085	0.084	0.082	0.079	0.075	0.073
ERS-P_T	0.084	0.064	0.026	0.006	0.001	0.000	0.000
NP-MZ_α^{GLS}	**0.100***	0.076	0.032	0.008	0.001	0.000	0.000
NP-MSB_α^{GLS}	0.088	0.069	0.031	0.008	0.002	0.000	0.000
NP-MZ_t^{GLS}	0.092	0.069	0.028	0.007	0.001	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.044	0.035	0.015	0.004	0.001	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.036	0.036	0.029	0.019	0.010	0.004	0.001
van der Waerden	0.037	0.039	0.047	0.059	0.079	**0.106***	**0.139***
Laplace	0.042	0.044	0.049	0.059	0.073	**0.091**	**0.111**
Wilcoxon	0.038	0.041	0.048	0.061	0.080	**0.105***	**0.139***
$\rho = 0.95$							
Dickey-Fuller	0.094	0.094	0.091	0.091	0.088	0.086	0.083
ERS-P_T	**0.141**	0.080	0.015	0.001	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.162***	**0.097***	0.022	0.002	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.139**	0.087	0.021	0.003	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.151**	0.089	0.019	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.074	0.048	0.011	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.047	0.047	0.044	0.034	0.021	0.010	0.004
van der Waerden	0.019	0.025	0.045	0.080	**0.139***	**0.221***	**0.326***
Laplace	0.028	0.035	0.049	0.076	**0.113**	**0.166**	**0.230**
Wilcoxon	0.020	0.027	0.047	0.083	**0.139***	**0.220***	**0.321***

Table 4: $n = 50$; $25,000$ replications; $f = \text{DE}$

Test	a						
	0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.078	0.078	0.078	0.078	0.078	0.078	0.078
ERS-P_T	0.045	0.045	0.045	0.045	0.045	0.045	0.045
NP-MZ_α^{GLS}	0.056	0.056	0.056	0.056	0.056	0.056	0.056
NP-MSB_α^{GLS}	0.052	0.052	0.052	0.052	0.052	0.052	0.052
NP-MZ_t^{GLS}	0.051	0.051	0.051	0.051	0.051	0.051	0.051
EM-$\hat{Q}^\mu(10,1)$	0.025	0.025	0.025	0.025	0.025	0.025	0.025
EM-$\hat{Q}^\mu(10,3.8)$	0.027	0.027	0.027	0.027	0.027	0.027	0.027
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.079	0.080	0.079	0.078	0.077	0.077	0.077
ERS-P_T	0.058	0.051	0.035	0.019	0.009	0.003	0.001
NP-MZ_α^{GLS}	0.070	0.062	0.045	0.025	0.012	0.004	0.001
NP-MSB_α^{GLS}	0.063	0.057	0.042	0.025	0.013	0.004	0.001
NP-MZ_t^{GLS}	0.063	0.056	0.040	0.023	0.011	0.004	0.001
EM-$\hat{Q}^\mu(10,1)$	0.030	0.027	0.018	0.011	0.005	0.002	0.001
EM-$\hat{Q}^\mu(10,3.8)$	0.031	0.028	0.023	0.018	0.012	0.007	0.003
van der Waerden	0.048	0.049	0.050	0.052	0.055	0.060	0.065
Laplace	0.049	0.050	0.052	0.054	0.059	0.064	0.070
Wilcoxon	0.048	0.049	0.050	0.053	0.058	0.064	0.069
$\rho = 0.975$							
Dickey-Fuller	0.083	0.082	0.081	0.078	0.076	0.072	0.070
ERS-P_T	0.081	0.058	0.025	0.007	0.001	0.000	0.000
NP-MZ_α^{GLS}	**0.097***	0.074	0.033	0.009	0.001	0.000	0.000
NP-MSB_α^{GLS}	0.084	0.068	0.032	0.009	0.001	0.000	0.000
NP-MZ_t^{GLS}	0.088	0.068	0.030	0.008	0.001	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.043	0.031	0.015	0.004	0.001	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.037	0.033	0.027	0.018	0.009	0.004	0.002
van der Waerden	0.037	0.042	0.053	0.071	**0.098**	**0.132**	**0.176**
Laplace	0.044	0.049	0.062	0.082	**0.111***	**0.151***	**0.197***
Wilcoxon	0.039	0.044	0.055	0.076	**0.105**	**0.145**	**0.195***
$\rho = 0.95$							
Dickey-Fuller	0.090	0.091	0.088	0.086	0.084	0.081	0.080
ERS-P_T	**0.134**	0.076	0.016	0.001	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.156***	**0.097***	0.023	0.002	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.135**	0.086	0.022	0.002	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.144**	0.088	0.020	0.002	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.071	0.042	0.011	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.046	0.045	0.043	0.032	0.019	0.009	0.005
van der Waerden	0.019	0.027	0.054	**0.105**	**0.186**	**0.294**	**0.420**
Laplace	0.032	0.044	0.076	**0.131***	**0.213***	**0.311**	**0.421**
Wilcoxon	0.021	0.029	0.061	**0.120**	**0.211***	**0.328***	**0.460***

Table 5: $n = 50$; $25,000$ replications; f Cauchy

Test	a 0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.077	0.077	0.077	0.077	0.077	0.077	0.077
ERS-P_T	0.025	0.025	0.025	0.025	0.025	0.025	0.025
NP-MZ_α^{GLS}	0.035	0.035	0.035	0.035	0.035	0.035	0.035
NP-MSB_α^{GLS}	0.048	0.048	0.048	0.048	0.048	0.048	0.048
NP-MZ_t^{GLS}	0.029	0.029	0.029	0.029	0.029	0.029	0.029
EM-$\hat{Q}^\mu(10,1)$	0.014	0.014	0.014	0.014	0.014	0.014	0.014
EM-$\hat{Q}^\mu(10,3.8)$	0.049	0.049	0.049	0.049	0.049	0.049	0.049
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.078	0.078	0.078	0.078	0.078	0.078	0.078
ERS-P_T	0.033	0.031	0.029	0.025	0.022	0.019	0.015
NP-MZ_α^{GLS}	0.042	0.042	0.038	0.033	0.027	0.023	0.019
NP-MSB_α^{GLS}	0.056	0.055	0.051	0.045	0.039	0.033	0.029
NP-MZ_t^{GLS}	0.036	0.036	0.033	0.028	0.024	0.020	0.016
EM-$\hat{Q}^\mu(10,1)$	0.017	0.018	0.015	0.013	0.011	0.009	0.007
EM-$\hat{Q}^\mu(10,3.8)$	0.047	0.046	0.042	0.039	0.036	0.033	0.030
van der Waerden	**0.149**	**0.150**	**0.151**	**0.155**	**0.159**	**0.164**	**0.170**
Laplace	**0.178***	**0.178***	**0.181***	**0.185***	**0.189***	**0.198***	**0.209***
Wilcoxon	**0.167**	**0.167**	**0.170**	**0.174**	**0.180**	**0.187**	**0.196**
$\rho = 0.975$							
Dickey-Fuller	0.080	0.080	0.079	0.079	0.078	0.077	0.076
ERS-P_T	0.046	0.042	0.034	0.027	0.020	0.016	0.013
NP-MZ_α^{GLS}	0.056	0.054	0.043	0.034	0.025	0.020	0.015
NP-MSB_α^{GLS}	0.070	0.066	0.056	0.045	0.035	0.030	0.023
NP-MZ_t^{GLS}	0.049	0.048	0.038	0.030	0.022	0.018	0.013
EM-$\hat{Q}^\mu(10,1)$	0.023	0.021	0.018	0.013	0.010	0.007	0.005
EM-$\hat{Q}^\mu(10,3.8)$	0.028	0.028	0.028	0.028	0.027	0.024	0.023
van der Waerden	**0.201**	**0.204**	**0.215**	**0.236**	**0.263**	**0.299**	**0.338**
Laplace	**0.254***	**0.259***	**0.275***	**0.299***	**0.333***	**0.376***	**0.422***
Wilcoxon	**0.233**	**0.238**	**0.253**	**0.277**	**0.308**	**0.348**	**0.394**
$\rho = 0.95$							
Dickey-Fuller	0.084	0.084	0.082	0.080	0.080	0.079	0.078
ERS-P_T	0.074	0.064	0.049	0.035	0.024	0.019	0.014
NP-MZ_α^{GLS}	0.090	0.080	0.058	0.040	0.029	0.021	0.016
NP-MSB_α^{GLS}	**0.100**	0.091	0.070	0.051	0.039	0.031	0.025
NP-MZ_t^{GLS}	0.082	0.072	0.052	0.036	0.025	0.019	0.014
EM-$\hat{Q}^\mu(10,1)$	0.038	0.032	0.023	0.016	0.011	0.007	0.005
EM-$\hat{Q}^\mu(10,3.8)$	0.026	0.026	0.028	0.030	0.030	0.028	0.026
van der Waerden	**0.225**	**0.232**	**0.261**	**0.305**	**0.366**	**0.434**	**0.507**
Laplace	**0.301***	**0.311***	**0.343***	**0.394***	**0.455***	**0.521***	**0.584***
Wilcoxon	**0.270**	**0.281**	**0.311**	**0.363**	**0.430**	**0.502**	**0.577**

Table 6: $n = 50$; $25,000$ replications; f skew-Normal (shape-parameter -10, mean 0, variance 1)

Test	a 0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.076	0.076	0.076	0.076	0.076	0.076	0.076
ERS-P_T	0.047	0.047	0.047	0.047	0.047	0.047	0.047
NP-MZ_α^{GLS}	0.056	0.056	0.056	0.056	0.056	0.056	0.056
NP-MSB_α^{GLS}	0.052	0.052	0.052	0.052	0.052	0.052	0.052
NP-MZ_t^{GLS}	0.051	0.051	0.051	0.051	0.051	0.051	0.051
EM-$\hat{Q}^\mu(10,1)$	0.024	0.024	0.024	0.024	0.024	0.024	0.024
EM-$\hat{Q}^\mu(10,3.8)$	0.025	0.025	0.025	0.025	0.025	0.025	0.025
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.075	0.077	0.077	0.078	0.078	0.078	0.077
ERS-P_T	0.060	0.052	0.035	0.016	0.007	0.002	0.001
NP-MZ_α^{GLS}	0.070	0.062	0.042	0.026	0.011	0.003	0.001
NP-MSB_α^{GLS}	0.064	0.058	0.042	0.026	0.011	0.003	0.001
NP-MZ_t^{GLS}	0.063	0.056	0.039	0.022	0.009	0.002	0.000
EM-$\hat{Q}^\mu(10,1)$	0.030	0.025	0.019	0.009	0.004	0.001	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.029	0.029	0.027	0.020	0.013	0.007	0.004
van der Waerden	0.049	0.049	0.049	0.051	0.054	0.058	0.064
Laplace	0.050	0.050	0.050	0.052	0.053	0.055	0.058
Wilcoxon	0.049	0.047	0.049	0.050	0.055	0.059	0.063
$\rho = 0.975$							
Dickey-Fuller	0.079	0.081	0.081	0.081	0.082	0.082	0.078
ERS-P_T	0.083	0.060	0.021	0.004	0.001	0.000	0.000
NP-MZ_α^{GLS}	**0.097**	0.072	0.032	0.006	0.001	0.000	0.000
NP-MSB_α^{GLS}	0.084	0.067	0.032	0.007	0.001	0.000	0.000
NP-MZ_t^{GLS}	0.089	0.066	0.028	0.005	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.042	0.030	0.012	0.002	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.034	0.034	0.032	0.021	0.012	0.005	0.001
van der Waerden	0.038	0.038	0.047	0.064	0.089	**0.125***	**0.167***
Laplace	0.043	0.045	0.049	0.057	0.070	**0.085**	**0.106**
Wilcoxon	0.038	0.039	0.048	0.064	0.087	**0.116**	**0.154**
$\rho = 0.95$							
Dickey-Fuller	0.087	0.090	0.093	0.095	0.095	0.092	0.093
ERS-P_T	**0.136**	0.074	0.011	0.001	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.157***	0.095	0.019	0.001	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.135**	0.088	0.020	0.001	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.146**	0.087	0.016	0.000	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.070	0.041	0.008	0.000	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.043	0.046	0.048	0.040	0.025	0.013	0.006
van der Waerden	0.019	0.021	0.043	0.087	**0.164***	**0.274***	**0.406***
Laplace	0.031	0.033	0.046	0.074	**0.111**	**0.163**	**0.231**
Wilcoxon	0.021	0.024	0.044	0.087	**0.156**	**0.256**	**0.379**

Table 7: $n = 100$; $25,000$ replications; $f = \mathcal{N}(0,1)$

Test	a						
	0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.063	0.063	0.063	0.063	0.063	0.063	0.063
ERS-P_T	0.047	0.047	0.047	0.047	0.047	0.047	0.047
NP-MZ_α^{GLS}	0.058	0.058	0.058	0.058	0.058	0.058	0.058
NP-MSB_α^{GLS}	0.052	0.052	0.052	0.052	0.052	0.052	0.052
NP-MZ_t^{GLS}	0.054	0.054	0.054	0.054	0.054	0.054	0.054
EM-$\hat{Q}^\mu(10,1)$	0.037	0.037	0.037	0.037	0.037	0.037	0.037
EM-$\hat{Q}^\mu(10,3.8)$	0.036	0.036	0.036	0.036	0.036	0.036	0.036
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.064	0.065	0.064	0.063	0.063	0.061	0.060
ERS-P_T	0.078	0.064	0.032	0.010	0.002	0.000	0.000
NP-MZ_α^{GLS}	0.091	0.073	0.039	0.013	0.003	0.000	0.000
NP-MSB_α^{GLS}	0.081	0.068	0.036	0.013	0.003	0.000	0.000
NP-MZ_t^{GLS}	0.086	0.068	0.035	0.011	0.002	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.056	0.046	0.026	0.009	0.002	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.052	0.045	0.033	0.018	0.009	0.004	0.001
van der Waerden	0.039	0.042	0.047	0.057	0.071	0.088	**0.110***
Laplace	0.044	0.046	0.050	0.056	0.065	0.076	0.090
Wilcoxon	0.040	0.042	0.047	0.055	0.069	0.085	**0.104***
$\rho = 0.975$							
Dickey-Fuller	0.075	0.075	0.075	0.073	0.070	0.069	0.068
ERS-P_T	**0.143**	0.085	0.016	0.001	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.165***	**0.100***	0.023	0.001	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.146**	0.090	0.022	0.001	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.156**	0.092	0.021	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	**0.103**	0.067	0.019	0.002	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.073	0.066	0.052	0.032	0.016	0.007	0.002
van der Waerden	0.018	0.023	0.044	0.083	**0.142***	**0.228***	**0.330***
Laplace	0.030	0.032	0.047	0.075	**0.113**	**0.163**	**0.228**
Wilcoxon	0.019	0.024	0.043	0.079	**0.138***	**0.214**	**0.314**
$\rho = 0.95$							
Dickey-Fuller	**0.104**	**0.104**	**0.108**	**0.113***	**0.126**	**0.142**	**0.164**
ERS-P_T	**0.319**	**0.132**	0.007	0.000	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.355***	**0.162***	0.014	0.000	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.311**	**0.145**	0.013	0.000	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.339**	**0.151**	0.012	0.000	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	**0.223**	**0.131**	0.023	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	**0.123**	**0.125**	**0.122***	**0.110***	0.090	0.068	0.048
van der Waerden	0.002	0.006	0.022	0.067	**0.162***	**0.311***	**0.501***
Laplace	0.012	0.019	0.036	0.071	**0.126**	**0.208**	**0.311**
Wilcoxon	0.003	0.007	0.022	0.065	**0.154**	**0.291**	**0.466**

Table 8: $n = 100$; $25,000$ replications; $f =$ DE

Test	a						
	0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.063	0.063	0.063	0.063	0.063	0.063	0.063
ERS-P_T	0.050	0.050	0.050	0.050	0.050	0.050	0.050
NP-MZ_α^{GLS}	0.058	0.058	0.058	0.058	0.058	0.058	0.058
NP-MSB_α^{GLS}	0.053	0.053	0.053	0.053	0.053	0.053	0.053
NP-MZ_t^{GLS}	0.053	0.053	0.053	0.053	0.053	0.053	0.053
EM-$\hat{Q}^\mu(10, 1)$	0.036	0.036	0.036	0.036	0.036	0.036	0.036
EM-$\hat{Q}^\mu(10, 3.8)$	0.037	0.037	0.037	0.037	0.037	0.037	0.037
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.066	0.066	0.065	0.063	0.063	0.062	0.062
ERS-P_T	0.080	0.060	0.029	0.009	0.002	0.000	0.000
NP-MZ_α^{GLS}	0.092	0.070	0.035	0.012	0.003	0.000	0.000
NP-MSB_α^{GLS}	0.083	0.065	0.033	0.013	0.003	0.000	0.000
NP-MZ_t^{GLS}	0.085	0.064	0.032	0.011	0.003	0.000	0.000
EM-$\hat{Q}^\mu(10, 1)$	0.058	0.046	0.023	0.007	0.002	0.000	0.000
EM-$\hat{Q}^\mu(10, 3.8)$	0.051	0.047	0.034	0.020	0.009	0.003	0.001
van der Waerden	0.040	0.043	0.051	0.062	0.080	**0.101**	**0.128**
Laplace	0.045	0.049	0.058	0.074	0.095	**0.123***	**0.158***
Wilcoxon	0.041	0.044	0.051	0.066	0.084	**0.108**	**0.140**
$\rho = 0.975$							
Dickey-Fuller	0.075	0.075	0.075	0.072	0.071	0.069	0.068
ERS-P_T	**0.145**	0.081	0.015	0.001	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.163***	0.094	0.021	0.002	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.144**	0.086	0.020	0.002	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.154**	0.087	0.018	0.002	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10, 1)$	**0.105**	0.065	0.016	0.002	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10, 3.8)$	0.070	0.066	0.052	0.033	0.016	0.007	0.003
van der Waerden	0.018	0.024	0.050	**0.101**	**0.182**	**0.292**	**0.426**
Laplace	0.030	0.043	0.080	**0.147***	**0.237***	**0.354***	**0.483***
Wilcoxon	0.019	0.028	0.058	**0.114**	**0.208**	**0.331**	**0.472**
$\rho = 0.95$							
Dickey-Fuller	**0.102**	**0.102**	**0.106**	**0.113**	**0.126**	**0.144**	**0.168**
ERS-P_T	**0.317**	**0.127**	0.008	0.000	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.348***	**0.153***	0.013	0.000	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.305**	**0.138**	0.013	0.000	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.332**	**0.144**	0.012	0.000	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10, 1)$	**0.217**	**0.124**	0.020	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10, 3.8)$	**0.120**	**0.123**	**0.121***	**0.107**	0.088	0.068	0.048
van der Waerden	0.003	0.007	0.029	**0.093**	**0.222**	**0.419**	**0.627**
Laplace	0.018	0.033	0.082	**0.170***	**0.299***	**0.448**	**0.599**
Wilcoxon	0.004	0.010	0.040	**0.119**	**0.270**	**0.474***	**0.677***

Table 9: $n = 100$; $25,000$ replications; f Cauchy

Test	a						
	0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.067	0.067	0.067	0.067	0.067	0.067	0.067
ERS-P_T	0.026	0.026	0.026	0.026	0.026	0.026	0.026
NP-MZ_α^{GLS}	0.033	0.033	0.033	0.033	0.033	0.033	0.033
NP-MSB_α^{GLS}	0.048	0.048	0.048	0.048	0.048	0.048	0.048
NP-MZ_t^{GLS}	0.028	0.028	0.028	0.028	0.028	0.028	0.028
EM-$\hat{Q}^\mu(10, 1)$	0.019	0.019	0.019	0.019	0.019	0.019	0.019
EM-$\hat{Q}^\mu(10, 3.8)$	0.053	0.053	0.053	0.053	0.053	0.053	0.053
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.069	0.070	0.070	0.070	0.070	0.071	0.071
ERS-P_T	0.041	0.041	0.038	0.035	0.030	0.027	0.023
NP-MZ_α^{GLS}	0.051	0.051	0.047	0.041	0.034	0.028	0.025
NP-MSB_α^{GLS}	0.066	0.066	0.061	0.053	0.046	0.040	0.035
NP-MZ_t^{GLS}	0.044	0.045	0.041	0.036	0.029	0.025	0.022
EM-$\hat{Q}^\mu(10, 1)$	0.030	0.029	0.026	0.023	0.020	0.017	0.014
EM-$\hat{Q}^\mu(10, 3.8)$	0.045	0.045	0.042	0.039	0.037	0.035	0.033
van der Waerden	**0.277**	**0.280**	**0.286**	**0.296**	**0.313**	**0.334**	**0.359**
Laplace	**0.339***	**0.341***	**0.351***	**0.366***	**0.387***	**0.412***	**0.442***
Wilcoxon	**0.314**	**0.319**	**0.326**	**0.339**	**0.357**	**0.381**	**0.410**
$\rho = 0.975$							
Dickey-Fuller	0.074	0.074	0.074	0.074	0.074	0.073	0.073
ERS-P_T	0.078	0.074	0.062	0.051	0.041	0.032	0.026
NP-MZ_α^{GLS}	0.092	0.086	0.072	0.057	0.046	0.036	0.029
NP-MSB_α^{GLS}	**0.104**	**0.098**	0.083	0.068	0.055	0.046	0.038
NP-MZ_t^{GLS}	0.083	0.078	0.064	0.050	0.040	0.032	0.026
EM-$\hat{Q}^\mu(10, 1)$	0.055	0.050	0.040	0.033	0.025	0.020	0.015
EM-$\hat{Q}^\mu(10, 3.8)$	0.042	0.041	0.040	0.039	0.038	0.036	0.034
van der Waerden	**0.353**	**0.360**	**0.381**	**0.414**	**0.462**	**0.517**	**0.576**
Laplace	**0.437***	**0.443***	**0.466***	**0.502***	**0.549***	**0.597***	**0.647**
Wilcoxon	**0.404**	**0.413**	**0.438**	**0.475**	**0.525**	**0.580**	**0.638**
$\rho = 0.95$							
Dickey-Fuller	0.082	0.082	0.080	0.081	0.083	0.084	0.085
ERS-P_T	**0.197**	**0.177**	**0.141**	**0.110**	0.086	0.069	0.057
NP-MZ_α^{GLS}	**0.218**	**0.196**	**0.155**	**0.117**	0.092	0.074	0.060
NP-MSB_α^{GLS}	**0.208**	**0.189**	**0.151**	**0.117**	0.091	0.075	0.061
NP-MZ_t^{GLS}	**0.207**	**0.184**	**0.146**	**0.109**	0.087	0.070	0.056
EM-$\hat{Q}^\mu(10, 1)$	**0.123**	**0.112**	0.087	0.065	0.050	0.039	0.030
EM-$\hat{Q}^\mu(10, 3.8)$	0.066	0.065	0.067	0.068	0.069	0.069	0.068
van der Waerden	**0.382**	**0.389**	**0.413**	**0.453**	**0.505**	**0.564**	**0.620**
Laplace	**0.460***	**0.470***	**0.494***	**0.528***	**0.566***	**0.608**	**0.646**
Wilcoxon	**0.440**	**0.448**	**0.478**	**0.517**	**0.565***	**0.618***	**0.664***

Table 10: $n = 100$; $25,000$ replications; f skew-normal (shape-parameter -10, mean 0, variance 1)

Test	a						
	0	1	2	3	4	5	6
$\rho = 1$							
Dickey-Fuller	0.063	0.063	0.063	0.063	0.063	0.063	0.063
ERS-P_T	0.051	0.051	0.051	0.051	0.051	0.051	0.051
NP-MZ_α^{GLS}	0.060	0.060	0.060	0.060	0.060	0.060	0.060
NP-MSB_α^{GLS}	0.053	0.053	0.053	0.053	0.053	0.053	0.053
NP-MZ_t^{GLS}	0.056	0.056	0.056	0.056	0.056	0.056	0.056
EM-$\hat{Q}^\mu(10,1)$	0.037	0.037	0.037	0.037	0.037	0.037	0.037
EM-$\hat{Q}^\mu(10,3.8)$	0.037	0.037	0.037	0.037	0.037	0.037	0.037
van der Waerden	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Laplace	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wilcoxon	0.050	0.050	0.050	0.050	0.050	0.050	0.050
$\rho = 0.99$							
Dickey-Fuller	0.065	0.065	0.064	0.062	0.062	0.062	0.061
ERS-P_T	0.079	0.062	0.028	0.009	0.001	0.000	0.000
NP-MZ_α^{GLS}	0.092	0.071	0.034	0.010	0.002	0.000	0.000
NP-MSB_α^{GLS}	0.082	0.066	0.033	0.010	0.002	0.000	0.000
NP-MZ_t^{GLS}	0.087	0.066	0.031	0.008	0.002	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	0.058	0.046	0.023	0.007	0.001	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.050	0.049	0.036	0.022	0.010	0.004	0.001
van der Waerden	0.042	0.042	0.048	0.060	0.077	**0.100***	**0.131***
Laplace	0.044	0.044	0.047	0.053	0.061	0.071	0.082
Wilcoxon	0.043	0.043	0.049	0.058	0.073	**0.092***	**0.117**
$\rho = 0.975$							
Dickey-Fuller	0.075	0.074	0.074	0.074	0.074	0.073	0.071
ERS-P_T	**0.145**	0.080	0.014	0.000	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.166***	0.094	0.017	0.001	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.145**	0.086	0.017	0.001	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.154**	0.087	0.015	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	**0.104**	0.065	0.015	0.001	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	0.069	0.070	0.056	0.038	0.020	0.008	0.003
van der Waerden	0.018	0.021	0.044	**0.096***	**0.177***	**0.294***	**0.443***
Laplace	0.030	0.032	0.045	0.069	**0.104**	**0.155**	**0.219**
Wilcoxon	0.020	0.024	0.046	0.089	**0.158**	**0.260**	**0.389**
$\rho = 0.95$							
Dickey-Fuller	**0.105**	**0.105**	**0.109**	**0.118***	**0.129**	**0.146**	**0.167**
ERS-P_T	**0.324**	**0.125**	0.004	0.000	0.000	0.000	0.000
NP-MZ_α^{GLS}	**0.350***	**0.154***	0.008	0.000	0.000	0.000	0.000
NP-MSB_α^{GLS}	**0.307**	**0.141**	0.009	0.000	0.000	0.000	0.000
NP-MZ_t^{GLS}	**0.333**	**0.143**	0.007	0.000	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,1)$	**0.222**	**0.130**	0.016	0.000	0.000	0.000	0.000
EM-$\hat{Q}^\mu(10,3.8)$	**0.122**	**0.124**	**0.126***	**0.115***	**0.096**	0.075	0.055
van der Waerden	0.002	0.004	0.020	0.077	**0.202***	**0.408***	**0.635***
Laplace	0.013	0.019	0.036	0.071	**0.130**	**0.218**	**0.336**
Wilcoxon	0.003	0.006	0.023	0.078	**0.191**	**0.373**	**0.584**

for testing the hypothesis $\rho = 1$. Accordingly, different versions exist, depending on the regression equation to be considered. These versions are presented, for example, in Hamilton (1994, Table 17.1). Two Dickey-Fuller tests are suited for both models (1) and (2). One possibility is to regress Y_t on a constant term and Y_{t-1} (as in (24)). In Hamilton (1994, Table 17.1), the behavior of this test is summarized in Case 2 and Case 3; denote by DF_1 the resulting Dickey-Fuller statistic. Another possibility is to regress Y_t also on a linear time-trend; in Hamilton (1994, Table 17.1) this is called Case 4; denote by DF the resulting Dickey-Fuller statistic. It is well-documented, however, that DF_1 yields non-similar tests—see, for example, Bhargava (1986), Hylleberg and Mizon (1989), or Dios-Palomares and Roldan (2006). Therefore, we rather use DF.

Turning to Tables 3-10, the figures speak for themselves:

(a) (validity) Irrespective of series lengths, starting values and underlying densities, Dickey-Fuller significantly over-rejects. The ERS-test P_T and NP-M^{GLS} tests are close to the nominal level, except for the Cauchy case, under which they are severely biased. The EM-test $\hat{Q}^{\mu}(10, 1)$ is uniformly and severely biased, as well as the EM-test $\hat{Q}^{\mu}(10, 3.8)$ which, however, has a much better behavior under Cauchy densities. The rank tests, as expected, perfectly match the nominal level.

(b) (short series lengths) Although of econometric practical relevance, $n = 50$ in this context is a very short series length, for which only the ERS-P_T and NP-M^{GLS} tests have some power at $\rho = 0.95$ and small $Y_0 - m$ values. Rank-based tests, however, have power under large values of $Y_0 - m$, and spectacularly outperform all their competitors under Cauchy densities.

(c) (heavy-tailed densities) All "classical" techniques, and, particularly so, Dickey-Fuller, fail miserably under Cauchy densities, while all rank-based ones are doing extremely well. This is all the more remarkable as the scores (van der Waerden, Wilcoxon, Laplace) considered here are not adapted to a heavy-tailed context, and Cauchy scores (see Hallin, Swan, Verdebout, and Veredas, 2011) are likely to perform even better.

(d) (impact of the starting value) Roughly, the deviation of Y_0 from the stationary mean m has a negative impact on the power of ERS-P_T, NP-M^{GLS}, EM-$\hat{Q}^{\mu}(10, 1)$ and EM-$\hat{Q}^{\mu}(10, 3.8)$ tests, and a positive impact on the rank-based ones; Tables 7, 8 and 10, for $\rho = 0.95$, are quite typical in that respect. The two families of procedures thus nicely complement each other (the deviation $Y_0 - m$, of course, is unknown in practice).

4. Conclusions

The rank-based tests we are proposing for the unit root hypothesis offer all the usual advantages of rank-based tests: *distribution-freeness*, exact finite sample sizes, and robustness. Moreover, they are flexible and efficient, in the sense that a reference density g can be chosen, which is such that semiparametric efficiency is achieved under density g. That reference density g can even be estimated, without affecting the validity of the test. Moreover, choosing a Gaussian reference density guarantees that our tests (of the van der Waerden type) are, (under (a)-asymptotics), uniformly locally more powerful than Dickey-Fuller test.

In finite samples, our simulation study shows that rank-based tests outperform the traditionally used Dickey-Fuller test, as well as several more recent competitors, for a broad range of initial values. Efficiency gains are particularly large when the underlying innovation density has fat tails. Our rank-based procedures thus nicely complement existing techniques.

The present paper focusses on the simplest setting possible. In particular, we assume the underlying innovations of the process to be i.i.d. This is needed in order to define optimality of testing procedures. However, extensions to models that allow for, e.g., parametric forms of heteroskedasticity are easily imagined.

A. Proofs

For ease of reference, we first provide a lemma on the joint convergence of a partial sum process and its rank-based version. Although based on existing results in the literature, this lemma as such does not seem to have been provided. The bottom line is that, where the partial sum process converges to a Brownian motion, its rank-based version converges to the Brownian bridge generated by that Brownian motion.

Lemma A.1 *Let $(U_1, \ldots, U_n)$ be i.i.d. standard uniformly distributed random variables and denote by R_t the rank of U_t. Let $\varphi : [0,1] \to \mathbb{R}$ be a measurable function satisfying $\int_0^1 \varphi(v)dv = 0$ and $\int_0^1 \varphi(v)^2 dv < \infty$. Define the partial sum processes $W_\varphi^{(n)}$ and $\widetilde{W}_\varphi^{(n)}$, both on $[0,1]$, by*

$$W_\varphi^{(n)}(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \varphi(U_t) \text{ and } \widetilde{W}_\varphi^{(n)}(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} E\{\varphi(U_t) | R_t\}. \tag{20}$$

Then, we have

$$\begin{bmatrix} W_\varphi^{(n)} \\ \widetilde{W}_\varphi^{(n)} \end{bmatrix} \Rightarrow \begin{bmatrix} W \\ \widetilde{W} \end{bmatrix}, \tag{21}$$

where W denotes a zero-drift Brownian motion with variance $\int_0^1 \varphi(v)^2 dv$ per unit of time and $\widetilde{W}$ its associated Brownian bridge: $\widetilde{W}(u) = W(u) - uW(1)$, $u \in [0,1]$. The convergence in (21) is on $D^2[0,1]$ equipped with the uniform topology.

Proof: It is well-known that weak convergence in $D^2[0,1]$ under the uniform topology follows from establishing convergence of marginals and asymptotic tightness, see, for example, Van der Vaart and Wellner (1993), Theorem 1.5.4.

Convergence of marginals for the partial sum process $W_\varphi^{(n)}$ is easily obtained from the central limit theorem. This implies also (joint) convergence of the marginals of its rank-based version $\widetilde{W}_\varphi^{(n)}$ using what is sometimes known as Hájek's representation theorem:

$$\widetilde{W}_\varphi^{(n)}(u) = W_\varphi^{(n)}(u) - uW_\varphi^{(n)}(1) + o_{\mathrm{P}}(1), \tag{22}$$

see Van der Vaart (2000), Theorem 13.5. In the notation of Van der Vaart (2000), we have $i = t$, $N = n$, $C_{Ni} = I\{t \leq un\}$, and $a_{Ni} = \mathrm{E}\{\varphi(U_t) | R_t = i\}$. From $\int_0^1 \varphi(v)\mathrm{d}v = 0$ we find $\bar{a}_N = 0$. Moreover, we have $\bar{c}_N = \lfloor un \rfloor / n \to u$.

Since marginal tightness implies joint tightness, the proof is concluded once we show that $\widetilde{W}_\varphi$ is tight in $D[0,1]$ under the uniform topology. This follows from Shorack and Wellner (1986). Take $c_{ni} = \mathrm{E}\{\varphi(U_t) | R_t = i\}$ and note that $\bar{c}_n = n^{-1} \sum_{i=1}^n c_{ni} = 0$, $n^{-1} \sum_{i=1}^n c_{ni}^2 \leq \int_0^1 \phi^2(u)du$. From this it easily follows that the conditions to Shorack and

Wellner (1986, Theorem 3.1) are satisfied: $\max_{i=1,\dots,n} c_{ni}^2/c_n^T c_n \to 0$ and $\bar{c}_n/\sqrt{c_{nn}^2} \to 0$. $\square$

Proof of Theorem 2.1: First recall that $g \in \mathcal{F}$ implies $\int_{u=0}^1 \varphi_g(u)\mathrm{d}u = 0$ and $\int_{u=0}^1 \varphi_g(u)^2\mathrm{d}u = I_g$, with $\varphi_g := -g'/g$. Moreover, under H_0, we have $\Delta Y_t = \mu + \epsilon_t$, so that the rank of ΔY_t amongst $\Delta Y_1, \dots, \Delta Y_n$ is the same as that of ϵ_t amongst $\epsilon_1, \dots, \epsilon_n$. Now, using $\widetilde{W}_{\varphi_g}^{(n)}$ as defined in Lemma A.1 and (10) with $U_t = F(\varepsilon_t)$, we obtain the asymptotic representation

$$T_g^{(n)} = \int_{u=0}^1 \left(u - \frac{1}{2}\right) \mathrm{d}\widetilde{W}_{\varphi_g}^{(n)}(u) + o_{\mathrm{P}}(1). \tag{23}$$

Lemma A.1 and the continuous mapping theorem thus imply that $T_g^{(n)}$ is asymptotically distributed as

$$\int_{u=0}^1 \left(u - \frac{1}{2}\right) \mathrm{d}\widetilde{W}(u) \sim N\left(0, I_g \int_{u=0}^1 \left(u - \frac{1}{2}\right)^2 \mathrm{d}u\right) = N\left(0, \frac{I_g}{12}\right). \qquad \square$$

Proof of Proposition 2.1: Case (ii) has been established in Jeganathan (1995, Section 7). For Case (i), the proof is analogous to that in Drost, Klaassen, and Werker (1997) for a pure location model. The rates of convergence obviously have to be adapted, as well as the form of the Fisher information matrix. Also, μ in our model (1) is a pure location parameter and its Fisher information, therefore, is I_f. The Fisher information for ρ is given by the limit of $n^{-3}\sum_{t=1}^n Y_{t-1}^2 \varepsilon_t^2$, analogously to the standard regression framework. Note that, under the null hypothesis ($\mu \neq 0, \rho = 1$), the drift μt in Y_t dominates the stochastic part, as

$$\frac{1}{n^3}\sum_{t=1}^n (Y_t - \mu t)^2 = \frac{1}{n^3}\sum_{t=1}^n \left(\sum_{s=1}^t \varepsilon_s\right)^2 \leq \frac{1}{n}\sum_{t=1}^n \left(\frac{1}{t}\sum_{s=1}^t \varepsilon_s\right)^2 \to 0 \text{ a.s.},$$

where the last convergence follows from a Cesàro mean argument and the strong law of large numbers. Consequently, we have

$$\lim_{n\to\infty} n^{-3}\sum_{t=1}^n Y_{t-1}^2 = \lim_{n\to\infty} n^{-3}\sum_{t=1}^n (\mu t)^2 = \mu^2/3 \ (a.s.),$$

which in turn leads to the Fisher information $\mu^2 I_f/3$ for ρ. $\square$

Proof of Theorem 2.2: The Hájek Asymptotic Representation result (23), combined with Lemma A.1, implies

$$T_g^{(n)} = \int_{u=0}^{1} \left(u - \frac{1}{2}\right) \mathrm{d}W_{\varphi_g}^{(n)}(u) + o_{\mathrm{P}}(1),$$

as

$$\int_{u=0}^{1} \left(u - \frac{1}{2}\right) \mathrm{d}\left[W_{\varphi_g}^{(n)} - \widetilde{W}_{\varphi_g}^{(n)}\right](u) \Rightarrow \int_{u=0}^{1} \left(u - \frac{1}{2}\right) \mathrm{d}\left[W - \widetilde{W}\right](u) = 0.$$

Also, Proposition 2.1 implies that

$$\log \mathrm{dP}_{(\mu,\rho_n);f}^{(n)}/\mathrm{dP}_{(\mu,1);f}^{(n)} = h\mu \int_{u=0}^{1} u\mathrm{d}W_{\varphi_f}^{(n)}(u) - h^2\mu^2 I_f/6 + o_{\mathrm{P}}(1).$$

As a result, the statistic $T_g^{(n)}$ and the log likelihood ratio are asymptotically jointly normal, with limiting covariance

$$h\mu I_{fg} \int_{u=0}^{1} u(u - 1/2)\mathrm{d}u = h\mu I_{fg}/12.$$

Le Cam's third lemma, see, e.g., Van der Vaart (2000), Section 6.7, now readily implies (16). $\square$

Proof of Corollary 2.1: The asymptotic distribution of the Dickey-Fuller test statistic is well-studied. For instance, it follows from Chapter 17 in Hamilton (1994) that, letting $\bar{Y}_n := n^{-1}\sum_{t=1}^{n} Y_{t-1}$,

$$\begin{aligned} n^{3/2}\left(\hat{\rho}_n^{DF} - 1\right) &= \frac{n^{-3/2}\sum_{t=1}^{n}\left(Y_{t-1} - \bar{Y}_n\right)\Delta Y_{t-1}}{n^{-3}\sum_{t=1}^{n}\left(Y_{t-1} - \bar{Y}_n\right)^2} \qquad (24) \\ &= n^{-1/2}\frac{12}{\mu}\sum_{t=1}^{n}\left(\frac{t}{n+1} - \frac{1}{2}\right)\varepsilon_t + o_{\mathrm{P}}(1). \end{aligned}$$

The null limiting distribution of $n^{3/2}\left(\hat{\rho}_n^{DF} - 1\right)$ thus is $\mathcal{N}(0, 12\sigma_f^2/\mu^2)$. As in Theorem 2.2, it follows from Le Cam's third lemma that its limiting distribution under the near (under (a)-asymptotics) unit root alternatives $H_1^{(n)}: \ \rho_n = 1 + hn^{-3/2}$ is $\mathcal{N}(h, 12\sigma_f^2/\mu^2)$, using the fact that $\mathrm{E}_f(-f'/f)(\varepsilon_t)\varepsilon_t = 1$. Incidentally, this shows that the least-squares estimator is (also) regular in this situation. $\square$

Acknowledgments

The authors gratefully acknowledge the constructive remarks by two anonymous referees, an Associate Editor and an Editor, which resulted in a major reorganization of the paper. They also thank Feike Drost, Alastair Hall, Ralph Koijen, Pierre Perron, and Per Mykland for useful comments.

The research by Marc Hallin was supported by Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, and by a Discovery Grant of the Australian Research Council.

References

References

Ahn, S.A., S.B. Fotopoulos, and L. He (2003). Unit root tests with infinite variance errors. *Econometric Reviews* 20, 461–483.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *The Review of Economic Studies* 53, 369–384.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

Breitung, J., and C. Gouriéroux (1997). Rank tests for unit roots. *Journal of Econometrics* 81, 7–27.

Callegari, F., N. Cappuccio, and D. Lubian (2003). Asymptotic inference in time series regressions with a unit root and infinite variance errors. *Journal of Statistical Planning and Inference* 116, 277–303.

Campbell, B., and J.M. Dufour (1995). Exact nonparametric orthogonality and random walk tests. *The Review of Economics and Statistics* 77, 1–16.

Campbell, B., and J.M. Dufour (1997). Exact nonparametric tests of orthogonality and random walk in the presence of a drift Parameter. *International Economic Review* 38, 151–173.

Chan, N.H., and C.Z. Wei (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367–401.

Chernoff, H., and I.R. Savage (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *The Annals of Mathematical Statistics* 29, 972–994.

Coudin, E., and J.M. Dufour (2009). Finite-sample distribution-free inference in linear median regression under heteroskedasticity and nonlinear dependence of unknown form. *Econometrics Journal* 12, 19–49.

Dickey, D.A., and W.A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit Root. *Journal of the American Statistical Association* 74, 427–431.

Dickey, D.A., and W.A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit Root. *Econometrica* 49, 1057–1072.

Dios-Palomares, R., and J.A. Roldan (2006). A strategy for testing the unit root in AR(1) model with intercept: a Monte Carlo experiment. *Journal of Statistical Planning and Inference* 136, 2685–2705.

Drost, F.C., C.A.J. Klaassen, and B.J.M. Werker (1997). Adaptive estimation in time-series models. *The Annals of Statistics* 25, 786–817.

Dufour, J.M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1389.

Elliott, G., and M. Jansson (2003). Testing for unit roots with stationary covariates. *Journal of Econometrics* 115, 75–89.

Elliott, G., and U.K. Müller (2006). Minimizing the impact of the initial condition on testing for unit roots. *Journal of Econometrics* 135, 285–310.

Elliott, G., T.J. Rothenberg, and J.H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Enders, W. (2004), *Applied Econometric Time Series* (2nd ed.). John Wiley & Sons.

Gushchin, A.A. (1996). Asymptotic optimality of parameter estimators under the LAQ Condition. *Theory of Probability and its Applications* 40, 261–272.

Haldrup, N. and M. Jansson (2006). Improving size and power in unit root testing. *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, ed. by T.C. Mills., and K. Patterson. New York: Palgrave Macmillan, 252-277.

Hallin, M., Y. Swan, T. Verdebout, and D. Veredas (2010). Rank-based testing in linear models with stable errors. *Journal of Nonparametric Statistics,* to appear.

Hamilton, J.D. (1994). *Time Series Analysis* (1st ed.). Princeton University Press.

Hasan, M.N. (2001). Rank tests of unit root hypothesis with infinite variance errors. *Journal of Econometrics* 104, 49–65.

Hasan, M.N., and R.W. Koenker (1997). Robust rank tests of the unit root hypothesis. *Econometrica* 65, 133–161.

Hylleberg, S., and G. Mizon (1989). A note on the distribution of the least squares estimator of a random walk with drift. *Economics Letters* 29, 225–230.

Im, K.S., M.H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115, 53–74.

Jansson, M. (2008). Semiparametric power envelopes for tests of the unit root hypothesis. *Econometrica* 76, 1103–1142.

Jansson, M., and M.J. Moreira (2008). Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74, 681–714.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Le Cam, L.M. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

Levin, A., C.-F. Lin, and C.-S. J. Chu (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–24.

Luger, R. (2003). Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity. *Journal of Econometrics* 115, 259–276.

Müller, U.K., and G. Elliott (2003). Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.

Ng, S., and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

Perron, P. (1988). Trends and random walks in macroeconomic time series. *Journal of Economics Dynamics and Control* 12, 297–332.

Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

Phillips, P.C.B., and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Ploberger, W. (2004). Complete class of tests when the likelihood is locally asymptotically quadratic. *Journal of Econometrics* 118, 67–94.

Ploberger, W. (2008). Admissible and nonadmissible tests in unit-root-like situations. *Econometric Theory* 24, 15–42.

Rachev, S.T., S. Mittnik, and J.-R. Kim (1998). Time series with unit roots and infinite-variance disturbances. *Applied Mathematics Letters* 11, 69–74.

Rothenberg, T.J., and J.H. Stock (1997). Inference in a nearly integrated autoregressive model with nonnormal innovations. *Journal of Econometrics* 80, 268–286.

Shorack, G., and J.A. Wellner (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Thompson, S.B. (2004). Optimal versus robust inference in nearly integrated non-Gaussian models. *Econometric Theory* 20, 23–55.

Thompson, S.B. (2004b). Robust tests of the unit root hypothesis should not be "modified". *Econometric Theory* 20, 360–381.

Van der Vaart, A. (2000). *Asymptotic Statistics* (first edition). Cambridge University Press.

Van der Vaart, A. and J. Wellner (1993). *Weak convergence and empirical processes* (2nd ed.). Springer-Verlag.

West, K.D. (1988). Asymptotic normality, when regressors have a unit root. *Econometrica* 56, 1397–1417.

White, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive Case. *The Annals of Mathematical Statistics* 29, 1188–1197.